7

The Normal Distribution and *z*-Scores

Use and Interpretation of the Standard Deviation

INTRODUCTION

The normal distribution provides the context within which we can most easily conceptualize the standard deviation as a measure of dispersion and understand its relationship to the mean. Technically, the normal distribution is what is termed a *sampling distribution*; as such, it will provide the basis for our interpretation of various inferential statistics in part III of this text. Although we are not yet in a position to explore the most important attributes of a normal distribution, some of its characteristics greatly facilitate description of data distributions and thus may be applied in reference to some concepts with which we have already become familiar. In this chapter, we will examine the normal distribution as a means of seeking greater understanding of the standard deviation. In later chapters, the theoretical basis of the distribution and its other uses will be accorded more extensive attention.

THE NORMAL DISTRIBUTION

Normal distributions are often referred to as *normal curves* or, because of their shape, *bell curves*. Normal distributions are symmetrical; they are identical on the two sides of the mean, with the mode falling at the mean and the number of cases at values farther from the mean becoming steadily fewer, as pictured in figure 7.1.

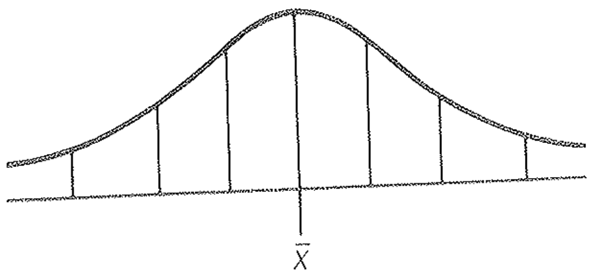

FIGURE 7.1 *A Normal Distribution Curve*

An interesting illustration of the tendency of many phenomena to distribute themselves normally is demonstrated in an exhibit at the Seattle Science Center. The exhibit is composed of a large frame box, only about a foot thick but several feet wide and several feet high. The two large sides are glass through which wood dividers may be seen that separate the inside of the box into a number of columns. Hundreds of metal balls are fed continuously into the box along a trough at the top that allows them to fall randomly into the various columns. As the demonstration proceeds, it quickly becomes apparent that most of the balls are falling into the center columns, with progressively fewer going into the outer columns. When all of the balls have fallen, their arrangement appears very much like the normal distribution curve in figure 7.1. Each time the demonstration is repeated, the distribution is slightly different, but an approximately normal distribution always results. Many real-life phenomena are known to be "normally distributed," with most cases grouped at the central tendencies and steadily fewer cases at values farther from the center. For instance, if data on intelligence scores or weights of adult males are collected for enough randomly selected cases, the scores will be distributed normally.

Characteristics of the Normal Curve

There are several constant properties of normal distributions and a few characteristics on which a distribution may vary and still be normal. As noted earlier, a normal distribution often is referred to as a normal curve because of the shape such distributions assume. Constant properties of the normal curve include the following:

1. A normal distribution is identical on both halves, with the mean, median, and mode falling at, and marking, the exact center of the distribution.
2. The areas under the normal curve are measured in terms of *standard deviation* units from the mean of the distribution.
3. All normal distributions have the same proportion of cases falling between specified standard deviation units.
4. From the peak, the curve falls steadily downward, first with a

slightly outward curve, then with an inward curve until it levels off.

5. The ends of the curve are not closed and do not touch the base.

PROPORTIONS UNDER THE NORMAL CURVE As indicated earlier, all normal distributions have the same proportion of cases falling between specified standard deviation units. Note that each standard deviation unit to the right of the mean is accompanied by a plus sign, and each to the left is accompanied by a minus sign. Units to the right are referred to as *above the mean* or as *plus* standard deviations, and units to the left are referred to as *below the mean* or *minus* standard deviation units.

Figure 7.2 indicates the proportions falling within each portion of a normal curve, with a range of three standard deviation units on each side of the mean. Converting these proportions to percentages, we can state that if a variable is normally distributed, 34.13% of the cases fall between the mean and one standard deviation above (to the right of) the mean; 13.59% of the cases fall between one and two standard deviations above the mean; 2.15% fall between two and three standard deviations above the mean, and only .13% fall more than three standard deviations above the mean. Although proportions may be calculated out to five standard deviations, the figures beyond three units become too small to deal with for most situations. Since the normal curve is symmetrical (identical on each side of the mean), the same percentages fall between the corresponding standard deviation units below (to the left of) the mean.

SUMMATIVE VALUES OF THE NORMAL CURVE Now look at the curve in figure 7.3. The values above the curve are summative values for one side of the curve. By summing the first two proportions to the right of the mean, we determine that 47.72% of all cases in a normal distribution will be between the mean and +2 standard deviations. Continuing to sum, we find that 49.87% of the normal curve is between the mean and three standard deviations above the mean. Since each side

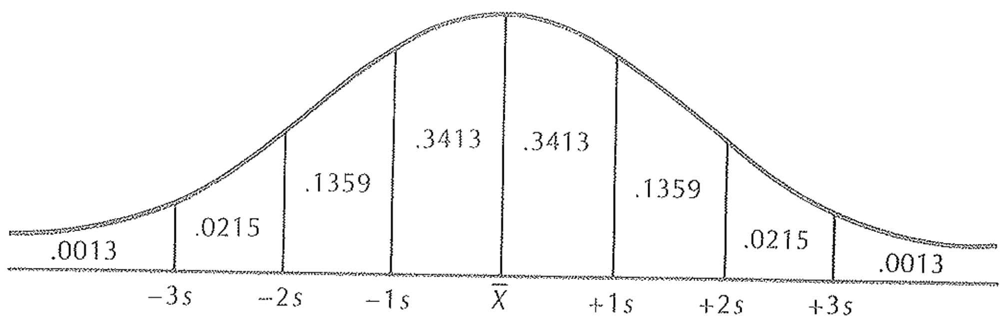

FIGURE 7.2 *Proportions of the Normal Curve*

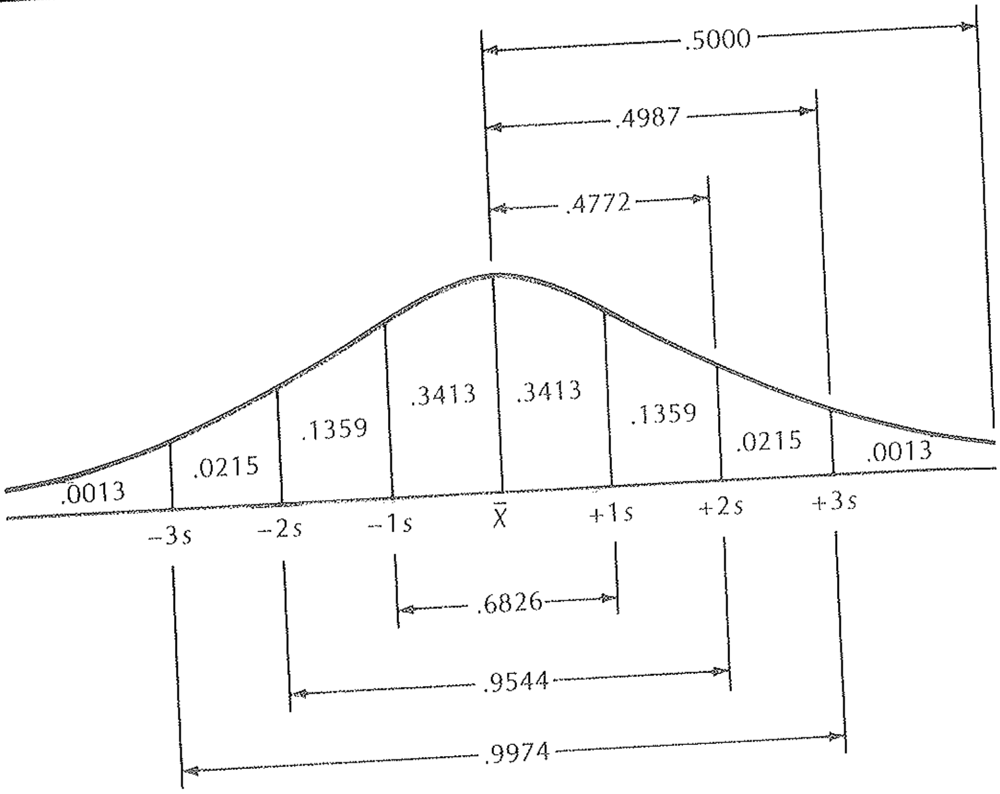

FIGURE 7.3
Summative Proportions of the Normal Curve

must equal half of the total area of the normal curve, a total of 50% of the curve is on each side of the mean.

Another way to look at the proportions within a normal distribution is presented below the curve in figure 7.3. By adding proportions on both sides of the mean, we determine that 68.26% of all cases will fall between ±1 (plus and minus 1) standard deviation; 95.44% will fall between ±2 standard deviations; and 99.44% will fall between ±3 standard deviations.

As indicated earlier, the ends of the normal curve remain open; the normal distribution theoretically extends into infinity in both directions. Measurements of the areas in the normal curve are available out to five standard deviation units, with five standard deviations marking .4999997133 of the curve. Because the proportions extend to infinity, no data set will perfectly fit the normal curve; however, many will be close approximations.

Application of the Areas of the Normal Curve

An understanding and appreciation of the meaning and use of the areas of the normal curve may be facilitated by application. As an example, consider the data used in chapters 5 and 6 for the fall semester GPAs of regularly admitted students, which were approximately normally distributed (see table 6.12). The curve on the left in figure 7.4 is labeled to reflect those data. Although all three central tendencies fall at the

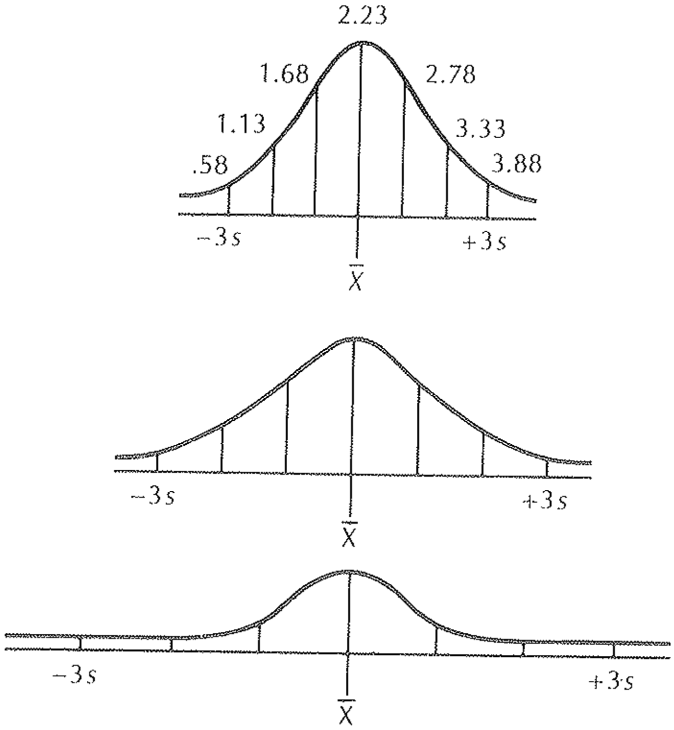

FIGURE 7.4
Variation in Normal Distribution Curves

exact center of the normal curve, the standard deviation is calculated using the mean (2.23) as the point of deviation reference, so that value is used to mark the center of the normal distribution curve. The base of the curve is measured with the standard deviation value (.55), subtracting the value of one standard deviation with each movement of one below the mean and adding one standard deviation for each unit above the mean. The value for $+1s$ is determined by adding s (in this case, .55) to the mean: $+1s = 2.23 + .55 = 2.78$. For $+2s$, sum the mean and two standard deviations: $2.23 + 2(.55) = 2.23 + 1.1 = 3.33$. Negative standard deviation values are derived by subtracting a corresponding number of s values from the mean. Thus, the GPA value corresponding to $-3s$ is $2.23 - 3(.55) = 2.23 - 1.65 = .58$. Given the aforementioned proportions, we could then expect approximately 34% of the cases to fall within one standard deviation on each side of the mean. From table 6.12, it can be determined that in fact 33.99% of the cases in the distribution fall between GPAs of 1.68 and 2.23, and 35.33% fall between 2.23 and 2.78.

As mentioned earlier, some characteristics of the normal distribution may vary. The variation occurs primarily in the density, or height of the distribution, and in its spread. For instance, all three curves in figure 7.4 are normal distributions, although they vary in height and width. The larger the standard deviation, the flatter the normal curve will be. For illustration, assume that each of the curves in figure 7.4 was constructed to reflect GPA distributions with a mean of 2.23 but

that each distribution had a different standard deviation. The curve on the left has the smallest standard deviation, so the cases are spread over a smaller range. The distribution on the right has the largest standard deviation, resulting in a flatter. more spread-out curve. Thus, whereas over 68% of the cases fall within a range of 1.1 GPA units in the curve on the left, 68% of the cases might be within a range of 1.8 GPA units in the curve on the right. Remember, though, that no matter what specific shape a distribution assumes, if it is normally distributed, a specified proportion of the cases will fall between the mean and each standard deviation unit.